Privacy, Data & De-identification

Acting Information Commissioner Timothy Pilgrim delivered this speech to CeBIT in Sydney on 2 May 2016.

Good afternoon.

I acknowledge the Wanngal people as the traditional custodians of this land, and pay my respect to elders past and present. I also thank CeBit for inviting me to speak to you today.

data is core to the development and delivery of most services, to paid and unpaid activities across the economy, and to better quality public policy

Today, I'm here to discuss privacy, data, de-identification; and the opportunities these present in an Australian context.

And the opportunities of bringing these three issues together in an integrated way are, I believe, significant – for Australian Government agencies and Australian businesses.

In fact, the raw potential that big data presents to both public and private sector alike is so extraordinary that it's a little hard to explain in words.

Yet a mathematician came close, in my view, and did so 202 years ago.

It's with a little trepidation that I refer to the works of great pioneering mathematicians in front of a CeBit audience, but those amongst you with a taste for the classics may recall "LaPlace's Demon".

This was Pierre-Simon LaPlace's famous treatise on determinism, which is often crudely summarised as the theory that if one could know the location and velocity of every object in the universe at a given point, one could predict the rest of history.

If that crude summation of LaPlace sounds suspiciously like history is throwing down a gauntlet to the power of big data analytics, then his actual words are even more prophetic:

> *We may regard the present state of the universe as the effect of its past and the cause of its future.*
>
> *If an intellect could know all forces that set nature in motion, and all positions of all items of which nature is composed, and* ***if this intellect were also vast enough to submit these data to analysis,*** *then nothing would be uncertain.*
>
> *The future, just like the past, would be present before its eyes.*[1]

Today we might simply quip that past metadata is the best predictor of future metadata.

But either way, the power of data-based prediction, which LaPlace could only theorise about, is now a reality.

Big data has changed the way we identify trends and challenges, as well as identify opportunities. As a result, it has the potential to bring about enormous social and economic benefits.

Trends drawn from big data can be used to personalise individuals' experiences, to target products and services, to improve health management, crime prevention, and emergency responses.

We've seen big data used not only to predict natural disasters, like flooding and earthquakes, but also to respond to them.

In 2015 the Humanitarian Data Exchange was used to help relief efforts following the Nepal earthquake. A task force of about 2,000 people from 80 countries analysed 'millions of Nepal-related tweets to build several databases'. This data helped produce quick-and-dirty maps to coordinate efforts by the government, the UN, and NGOs.[2]

And as the amount of data is growing exponentially, that potential can only increase.

As the Productivity Commission's recent Issues Paper explains, 5 billion gigabytes of data was the amount of data generated worldwide in the year 2002. We now generate it *every two days*.

And when I say "we", I mean it truly is a global community effort.

When we wake up, we check Twitter or Facebook or our emails.

Over breakfast we use our iPads to read the news.

Before work we might fit in a quick session at the gym – our Fitbit tracking our progress.

As we head off to work our smart phone pings the towers along the way.

Swiping our work pass we enter the building before logging onto our computers.

With each step we take we are, quite literally, creating more and more data – potentially revealing more and more about ourselves.

[1] *Pierre Simon Laplace, A Philosophical Essay on Probabilities*

[2] *How The Candy Crush Of Data Is Saving Lives In Nepal*

And as our digital touch points increase, and the Internet of Things becomes more and more embedded in our everyday lives, the data we create becomes increasingly valuable.

Valuable to both private and public sector alike.

The Prime Minister made this clear when he released the *Australian Government Public Data Policy Statement* at the end of last year. It recognises data held by the Australian Government as a strategic national resource that holds considerable value for growing the economy, improving service delivery and transforming policy outcomes for the Nation.

This priority is reflected in the fact that the Prime Minister's own department has established a Public Data Branch to lead data innovation across the public service.

After all, the policy and service delivery improvements that can be yielded if this national resource can be shared and built upon are immense.

Accordingly, the Productivity Commission has been tasked with looking at *Data Availability and Use*. In its Issues Paper, the Commission argues that data is core to the development and delivery of most services, to paid and unpaid activities across the economy, and to better quality public policy.

Both of these key Government papers *also* make it clear that upholding the highest standards of data security and privacy are *critical*. And I welcome this focus.

Because my Office, the Office of the Australian Information Commissioner, has long supported the view that public information is a national asset.

Indeed, the FOI Act, which we administer alongside the Privacy Act, explicitly describes government information as a national resource.

We understand that the potential of that resource may be best realised when data can be shared, used and built upon.

But we also understand, and hope is evident, that this can only occur sustainably, if privacy is integral to the equation.

Simply put, a successful data-driven economy needs a strong foundation in privacy.

Our experience and community research shows that by and large people do want their personal information to work for them, provided that they know it is working *for* them. When there is transparency in how personal information is used, it gives individuals choice and confidence that their privacy rights are being respected.

Accordingly, good privacy management and great innovation go hand in hand.

Because when people have confidence about how their information is managed, they are more likely to support the use of that information to provide better services.

In fact, their expectations often become entirely supportive.

Most people *do* expect organisations to use their information where it's necessary to provide them with the services they want or to improve on those services.

They do expect law enforcement agencies to use information resources to stop crime and to keep people safe.

However, people also want to know how their information is being used, who has access to it, and what that means for them in terms of their personal identity.

Accordingly, privacy law – often misunderstood to be about secrecy, is really underpinned by transparency and accountability.

Accordingly, good privacy management and great innovation go hand in hand

And by ensuring organisations are transparent and responsible when handling personal information, privacy management strengthens customer trust.

Building this trust is key to our big data challenges – whether sought in the form of customer confidence or political mandate.

As the Chairman of the Productivity Commission has said 'the significant evolution in data collection and analysis seen in recent times suggests that the culture, standards and policy structures that have been applied to big data analytics may need to move out of the back room and into the showroom if community confidence and wide opportunity for innovation are to be maximised.'

And I agree.

We know from my Office's longitudinal surveys into community attitudes to privacy, that Australians are becoming increasingly conscious of personal data issues.

The majority of Australians – 60 percent – have decided not to deal with an organisation due to concerns about how their personal information will be used. And significantly, 97 percent of Australians don't like their personal information to be used for a secondary purpose.

This is critical to big data. Because big data projects will often involve secondary use of data.

If that data finds its source in personal information, then we have a clear dissonance between our known and understandable desire that our personal information works *for us* and for the purposes *we* explicitly provided it for versus the demonstrable innovative power of that data to improve our services and lives.

Addressing this dissonance will require a multi-pronged approach.

Part of it will lie in making the case as to how, through secondary uses, our personal information is still clearly working for our benefit, either directly or communally – and numerous research fields point to the potential to make this case.

De-identification is a smart and contemporary response to the privacy challenges of big data

Part of it will lie in greater security and protection of the personal information – and a determined approach to counter would-be disrupters of our national data resource – as the Government's new *Cyber Security Strategy* reveals.

But part of the solution, and potentially a significant part I suggest, lies in getting de-identification right, and right such that government agencies, regulators, businesses and technology professionals have a common understanding as to what "getting it right" means.

At the moment, that common understanding is not evident.

I know, for example, that when I've previously spoken about precautions with *anonymised* data sets the result has been reporting that I'm advising to treat *de-identified* data as personal information.

This would be illogical advice at best, because correctly de-identified information is, by definition, no longer personal information.

To be clear, this has never been my Office's view, but the example highlights the current haze around this issue and the need to obtain an agreed understanding.

There may well be people in this room who are thinking, "well, I understand the distinction between anonymised and de-identified just fine thank you" and I'm sure that's true.

But as per the Productivity Commission's point, we need to move this knowledge out of the backroom and in to the showroom in order to build public confidence in this potential privacy solution.

Because it *is* a potential solution.

De-identification is a smart and contemporary response to the privacy challenges of big data – using the same technology that allows data analytics to strip data sets of their personal identification potential, while retaining the research utility of the data.

When done correctly, de-identified information is no longer personal information and is therefore outside the scope of the *Privacy Act*.

But what does "done correctly" entail?

De-identified means de-identified in *whose* hands?

And in *what* use?

If I am the collector of the personal information, am I obliged to have regard to the re-identification potential of data in its *current* context, the next *foreseeable* context, or *any* context?

And what about the ability of data analytics to create entirely new and personal information – raising the prospect of an entity effectively collecting new personal information by creating it?

These are all pertinent questions, but if you think I'm going to give clear and simple answers now, then I'm afraid you are in for disappointment.

This is for a good reason. Namely, the *Privacy Act* is principles, not prescription, based, and ultimate answers as to compliance with it will often be bespoke to the circumstances.

This is certainly true if your preferred solution to privacy governance is de-identification. The specific changes required to your data set will arrive as the result of a risk based assessment of the data's potential use, disclosure and re-identification prospects.

While the principles remain constant – and are already covered in our existing guidance on information sharing and de-identification – the solutions executed are often bespoke to the data and its intended use.

This is why it's not desirable to try and provide a prescriptive, template based, tick-a-box guide to de-identification.

It is why, despite already having guidance in this space, we will be opening up consultation on renewed guidance this year.

Because it is clear from the speed at which this big data is evolving that any privacy solution which is *purely* regulator-driven, without the voice of industry, consumers and government agencies to inform it, will not serve our purposes here.

To be clear, the *Privacy Act* principles, and the accountability of my office to regulate them, are both established, clear and ongoing – but ensuring that the application of these regulatory principles is as practical as possible in real world examples, is of benefit to both regulator and regulated alike.

This was the primary point of my recent, perhaps wistful, comparison between our current national race to harness the potential of big data, and the technological pioneering of the moon race.

As was the case with *that* great technological goal, potential solutions to balancing the democratic, strategic and commercial benefits of big data will lie in a multi-sector co-operation.

The OAIC understands that this is an area of regulation where agreed industry terms and standards will be critical – not only to the actual efficacy of de-iden-